The Capacity Loss of Uncorrelated Equi-Power Gaussian Input over MIMO Channel

Tal Philosof
Dept. of Electrical Engineering - Systems
Tel-Aviv University
Tel-Aviv 69978, ISRAEL
talp@eng.tau.ac.il

Ram Zamir
Dept. of Electrical Engineering - Systems
Tel-Aviv University
Tel-Aviv 69978, ISRAEL
zamir@eng.tau.ac.il

Abstract

We investigate the capacity loss of uncorrelated Gaussian input with equal power (i.i.d Gaussian input) over a multi-input multi-output linear additive noise (not necessarily Gaussian nor uncorrelated) channel. Previous work showed that this is the best input in the case of Gaussian noise, assuming the channel matrix is known at the receiver but unknown at the transmitter. We show that i.i.d Gaussian is a robust input also when the noise is not Gaussian and derive bounds for the capacity loss. Specifically, we show that for n_t transmit antennas and n_r receive antennas, the capacity loss of i.i.d Gaussian input is smaller than $\min\{\frac{n_t}{2}, \frac{n_r}{2}\log_2(1+\frac{n_t}{n_r})\}$ bits, for any noise and channel matrix. This bound is apparently not tight. Nevertheless, for the case of Gaussian noise we derive a stronger bound: $\frac{n_r}{2}\log_2(\frac{n_t}{n_r})$ bits for $1 \leq n_r \leq n_t/e$ and $\frac{n_t}{2}\frac{log_2(e)}{e}$ bits for $n_r \geq n_t/e$ which is tight for a "critical" channel matrix.

1 Introduction

Consider a multi-input multi-output (MIMO) channel model with additive noise

$$\mathbf{Y} = H\mathbf{X} + \mathbf{N}, \tag{1}$$

where $H \in \mathbb{R}^{n_r \times n_t}$ is the channel matrix, n_t and n_r are the number of transmit and receive antennas, respectively, $\mathbf{X} \in \mathbb{R}^{n_t}$ and $\mathbf{Y} \in \mathbb{R}^{n_r}$ are the transmitted and received symbols, respectively, while $\mathbf{N} \in \mathbb{R}^{n_r}$ is a general (not necessarily Gaussian nor independent) additive noise. The power constraint is $tr(R_{\mathbf{xx}}) = E||\mathbf{X}||^2 \leq P$, where tr is the trace operator and $R_{\mathbf{xx}} = E\mathbf{X}\mathbf{X}^{\mathbf{T}}$ is the covariance matrix of $\mathbf{X}$.

The capacity achieving input for this channel depends on both the channel matrix and the noise statistics. Therefore, it requires knowledge of these parameters not only at the receiver (which is relatively easy) but also at the transmitter. For the case of Gaussian noise, it was shown in [4] that the *compound channel* capacity [2], where the channel matrix belongs to "isotropic" set, is achieved using i.i.d Gaussian input. In [3], the capacity loss of i.i.d Gaussian input was considered for Rician MIMO channel with additive Gaussian noise.

A white Gaussian transmission over a single-input single-output (SISO) additive noise channel with power constraint, where the noise distribution is arbitrary, was considered in [6]. It was shown that the information rate loss due to using white Gaussian input instead of the optimum input distribution is bounded by half bit per channel use, while for colored Gaussian noise the loss for not performing the water filling optimization is at most $\log_2(e)/2e = 0.265$ bit per channel use.

In this paper we extend these results to i.i.d Gaussian input over MIMO channel with arbitrary noise and with Gaussian noise. Throughout this paper we use the term capacity loss for the gap between the capacity of MIMO channel with perfect channel state information (CSI) at the transmitter and receiver (which takes into account the noise statistics), and the mutual information achieved using i.i.d Gaussian input.

2 Main Result

2.1 Capacity Loss for Arbitrary Noise

The capacity with perfect CSI and general noise statistics is given by

$$C(H, \mathbf{N}, P) \triangleq \sup_{\mathbf{X}: E||\mathbf{X}||^2 \leq P} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}), \tag{2}$$

where the maximization is over all possible joint distributions of the input vector satisfying the power constraint. We present a uniform upper bound for the capacity loss due to using i.i.d Gaussian input over MIMO channel with arbitrary additive noise. Similarly to the concept used in [6] for the SISO channel case, we start by deriving the following lemma for a MIMO channel.

Lemma 1. *For a MIMO channel (1) with input $\mathbf{X}'$ and noise $\mathbf{N}$*

$$C(H, \mathbf{N}, P) - I(\mathbf{X}'; H\mathbf{X}' + \mathbf{N}) \leq C^*(H, \mathbf{X}'), \tag{3}$$

where

$$C^*(H, \mathbf{X}') \triangleq \sup_{\mathbf{Z}: E||\mathbf{Z}||^2 \leq P} I(H\mathbf{Z}; H\mathbf{Z} + H\mathbf{X}'). \tag{4}$$

Proof. The proof is a simple extension of the proof of [6, Lemma 1]. □

Note that the quantity $C^*(H, \mathbf{X}')$ is independent of the noise distribution $\mathbf{N}$, hence it bounds the capacity loss for using $\mathbf{X}'$ for any additive noise channel with channel matrix H. The worst capacity loss for using $\mathbf{X}'$ is bounded from above by the maximum of this quantity over the channel matrix $\sup_H C^*(H, \mathbf{X}')$, where H belongs to unrestricted channel matrix set. The smallest bound is achieved by $\mathbf{X}'$ that minimizes the maximum value of $C^*(H, \mathbf{X}')$, i.e, the $\mathbf{X}'$ which achieves

$$C^* \triangleq \inf_{\mathbf{X}': E||\mathbf{X}'||^2 \leq P} \sup_H C^*(H, \mathbf{X}'). \tag{5}$$

Lemma 2. *The minimum of the worst capacity loss bound (5) is achieved by $\mathbf{X}' = \mathbf{X}^*$, where $\mathbf{X}^* \sim \mathcal{N}(0, \frac{P}{n_t} I_{n_t})$ is i.i.d Gaussian distribution and I_{n_t} is the $n_t \times n_t$ dimensional identity matrix.*

Proof. The proof is given in Appendix A. □

As a consequence we have the following theorem.

Theorem 1. *(Arbitrary Noise) For the $n_r \times n_t$ MIMO channel (1) with any noise $\mathbf{N}$ and channel matrix H, the capacity loss is bounded by*

$$C(H, \mathbf{N}, P) - I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}) \leq C^*, \tag{6}$$

where $\mathbf{X}^* \sim \mathcal{N}(0, \frac{P}{n_t} I_{n_t})$ *is i.i.d Gaussian input,* $C(H, \mathbf{N}, P)$ *is the channel capacity, and*

$$C^* = \begin{cases} \frac{1}{2} n_t & , n_r \geq n_t \\ \frac{n_r}{2} \log_2 (1 + \frac{n_t}{n_r}) & , 1 \leq n_r \leq n_t, \end{cases} \tag{7}$$

bit per vector channel use, where n_r *and* n_t *are the number of receive and transmit antennas, respectively.*

Proof. The proof is given in Appendix B. □

For $n_r \geq n_t$, where there are more receive antennas than transmit antennas, the loss for using i.i.d Gaussian input is at most $\frac{1}{2}$ bit per channel use per transmit antenna, similarly to the result in [6]. However, for $1 \leq n_r < n_t$ when there are more transmit antennas, the bound is $\frac{n_r}{2} \log_2 (1 + \frac{n_t}{n_r})$ bits which is less than $\frac{n_t}{2}$ but more than $\frac{n_r}{2}$ bits. Therefore, we can not state that the loss is bounded by half bit per degree of freedom, i.e $\frac{1}{2} \min(n_t, n_r)$. For example, for one receive antenna the bound is $1/2 \log_2(1 + n_t)$ bit, which goes to infinity as the number of transmit antennas increases. In Figure 1, we illustrate the behavior of the bound on the rate loss for constant n_t. The function $\frac{n_r}{2} \log_2 (1 + \frac{n_t}{n_r})$ is concave w.r.t n_r (assuming n_r is a continues variable). We believe that

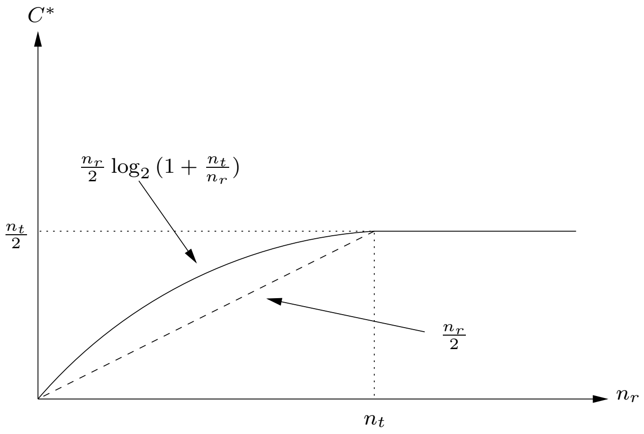

Figure 1: The Capacity Loss of a MIMO channel

this upper bound is not tight, and the distance from the true curve of the worst loss is still unknown. Nevertheless, for the case of Gaussian noise, we show in Section 2.2 a tighter bound which can actually be achieved, and can be considered as a lower bound for the loss for an arbitrary noise.

2.2 Capacity Loss for Gaussian Noise

The bound in Theorem 1 takes into account two effects. One is the "shaping loss" due to Gaussian input being mismatched to the higher order statistics of the noise, and the other is "water filling loss" due to white input being mismatched to the matrix H and to the noise covariance. In this section we focus on the second effect by restricting attention to Gaussian noise.

We define a critical channel matrix H^C which is a worst case channel matrix which has $\min\{n_r, n_t/e\}$ identical non-zero elements on the main diagonal and all off-diagonal elements are zero.

Theorem 2. *(Gaussian Noise) For a MIMO channel (1), with additive Gaussian noise* $\mathbf{N}$*,*

$$C(H, \mathbf{N}, P) - I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}) \leq C_G^*, \tag{8}$$

where $\mathbf{X}^* \sim \mathcal{N}(0, \frac{P}{n_t} I_{n_t})$ *is i.i.d Gaussian input, and*

$$C_G^* = \begin{cases} \frac{n_t}{2} \frac{\log_2(e)}{e} & , n_r \geq n_t/e \\ \frac{n_r}{2} \log_2\left(\frac{n_t}{n_r}\right) & , 1 \leq n_r \leq n_t/e. \end{cases} \tag{9}$$

Equality holds if $H = H^C$ *is the critical channel matrix defined above. Thus, for a Gaussian MIMO channel the capacity loss of i.i.d Gaussian input is at most* $\frac{\log_2(e)}{2e} \simeq 0.265$ *bit per channel use per transmit antenna.*

Proof. The proof of the bound is given in Appendix C. A case of equality is given below. □

The tightness of the bound can be shown for the case that $H = H^C$ with high SNR, and uncorrelated unit variance noise. For $1 \leq n_r \leq n_t/e$ the capacity is $C(H^C, \mathbf{N}, P) \simeq \frac{1}{2} n_r \log_2 \frac{P}{n_r}$, and the rate achieved using $\mathbf{X}^*$ is $I(\mathbf{X}^*; H^C\mathbf{X}^* + \mathbf{N}) \simeq \frac{1}{2} n_r \log_2 \frac{P}{n_t}$, so the rate loss is given by $C(H^C, \mathbf{N}, P) - I(\mathbf{X}^*; H^C\mathbf{X}^* + \mathbf{N}) \simeq \frac{1}{2} n_r \log_2 \frac{n_t}{n_r}$. For $n_r \geq n_t/e$ the channel matrix H^C has effectively only n_t/e sub-channels, therefore the capacity is $C(H^C, \mathbf{N}, P) \simeq \frac{n_t}{2e} \log_2 \frac{P}{n_t/e}$, and the rate achieved using $\mathbf{X}^*$ is $I(\mathbf{X}^*; H^C\mathbf{X}^* + \mathbf{N}) \simeq \frac{n_t}{2e} \log_2 \frac{P}{n_t}$, so the rate loss is given by $C(H^C, \mathbf{N}, P) - I(\mathbf{X}^*; H^C\mathbf{X}^* + \mathbf{N}) \simeq \frac{1}{2} n_t \frac{\log_2 e}{e}$.

It follows from Theorem 2 that the bound C_G^* on the capacity loss can be achieved. Figure 2 illustrates both bounds of Theorem 1 and Theorem 2. It can be seen that the bound for Gaussian noise is strictly less than that for arbitrary noise.

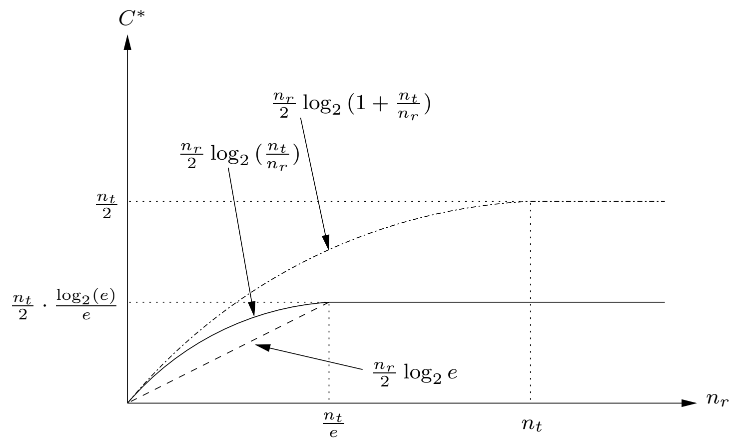

Figure 2: The two bounds on the capacity loss of i.i.d. Gaussian input

3 Discussion

We presented two uniform upper bounds for the capacity loss for using i.i.d Gaussian input over a MIMO channel with additive noise. The first bound is for arbitrary noise (not necessarily Gaussian) while the second one is for (colored) Gaussian noise. The former bound is apparently not tight. On the other hand, for the Gaussian noise case the bound on the loss can actually be achieved. Due to its tightness, the latter bound provides a lower bound for the maximum loss for general noise. For multiple transmit antennas and single receive antenna these upper and lower bounds are close, so we have a good approximation for the true capacity loss for arbitrary noise.

There are several scenarios in which Theorem 1 can be applied in communication over a MIMO channel where the channel matrix and noise statistics are known at the receiver. For instance, the receiver may have perfect knowledge of the channel matrix and the noise statistics by a learning process, while the transmitter may stay ignorant to the channel and the noise statistics. In this case, a limited feedback link is needed between the receiver and the transmitter in order to inform the encoder about the rate. Due to the transmitter ignorance to the channel matrix and the noise statistics, a capacity loss is incurred. For example if the channel matrix/noise statistics are changing too fast, it is worthwhile to use fixed shape codebook (such as good codebooks for an AWGN channel), since it is not possible to adapt the codebook frequently.

Another scenario is a system that uses "rateless" code, where the transmitter has fixed codebook while the receiver uses for decoding only part of the transmitted block which is needed for reliable decoding and ignores the rest of the block. This concept was introduced in [5] for transmission of common messages in broadcast channels. In this scheme there is no need for feedback link between the receiver and the transmitter, since the receiver determines the effective rate solely. Another case that does not require feedback link is of stochastic channel matrix with high order dimensions, i.e., $n_r n_t \gg 1$, where each element of the channel matrix is drawn i.i.d. According to the law of large numbers, the capacity (which is an average rate per received antenna) converges to a fixed value, which is therefore known to the transmitter (assuming that the transmitter knows the channel matrix and the noise statistics).

The i.i.d Gaussian input can generate a single codebook for all transmit antennas, or alternatively, a set of independent codebooks, one for each transmit antenna. In the former scheme the transmitter needs to know the MIMO channel capacity, while in the second scheme it requires to know the rate that each transmit antenna can carry. This equivalent to achievable rate point in capacity region of the corresponding multiple access channel (MAC) where the transmitters have independent codebooks. It follows that the capacity loss bound applies for the MAC.

We currently study the loss of using sub-optimal codebook for point to multi-point problems.

Appendix

A. Proof of Lemma 2

We first show that without loss of generality we can assume that the channel matrix H is a full rank $n_r \times n_t$ matrix with orthonormal rows where $n_r \leq n_t$, i.e, $rank(H) = n_r$. Generally, a channel matrix G is $n_r \times n_t$ real matrix with rank r not necessarily a full rank, thus $r \leq \min(n_t, n_r)$. Since the vectors $G\mathbf{X}^*$ and $G(\mathbf{Z}+\mathbf{X}^*)$ have at most r linearity independent components, therefore there is a matrix $T \in \mathbb{R}^{r \times n_r}$ such that

$$\begin{aligned} h(G(\mathbf{Z}+\mathbf{X}^*)) &= h(TG(\mathbf{Z}+\mathbf{X}^*)) && (10) \\ h(G\mathbf{X}^*) &= h(TG\mathbf{X}^*), && (11) \end{aligned}$$

hence

$$I(G\mathbf{Z}; G\mathbf{Z}+G\mathbf{X}^*) = I(TG\mathbf{Z}; TG\mathbf{Z}+TG\mathbf{X}^*). \quad (12)$$

The equivalent channel matrix $\tilde{G} = TG$ is $r \times n_t$ matrix, where $r \leq n_t$ with rank r. Using Gram Schmidt process $\tilde{G}$ can be written as $\tilde{G} = RH$, where $R \in \mathbb{R}^{r \times r}$ is non-singular lower triangular matrix and $H \in \mathbb{R}^{r\times n_t}$ is a full rank with orthonormal rows, thus

$$\begin{aligned} I(G\mathbf{Z}; G\mathbf{Z}+G\mathbf{X}^*) &= I(RH\mathbf{Z}; RH\mathbf{Z}+RH\mathbf{X}^*) \quad (13)\\ &= I(H\mathbf{Z}; H\mathbf{Z}+H\mathbf{X}^*), \quad (14)\end{aligned}$$

where (13) is from (12) and using that $TG = RH$, (14) is from the fact that multiplication by R which is non-singular square matrix does not change the mutual information. As a consequence from above, when we are dealing with the capacity loss term $I(H\mathbf{Z}; H\mathbf{Z}+H\mathbf{X}^*)$, the channel matrix set can be restricted to full rank channel matrices $H \in \mathbb{R}^{n_r\times n_t}$ with orthonormal rows where $n_r \le n_t$.

Initially, we show that

$$\frac{n_r}{2}\log_2(1+\frac{n_r}{n_t}) \le \sup_H C^*(H,\mathbf{X}'),\ \forall\ R\mathbf{x}'\mathbf{x}'. \quad (15)$$

For any $R_{\mathbf{x}'\mathbf{x}'}$ there is unitary matrix Q such that $R_{\mathbf{x}'\mathbf{x}'} = Q\Lambda_{\mathbf{x}'}Q^T$ where $\Lambda_{\mathbf{x}'}$ is $n_t \times n_t$ diagonal matrix with increasing order of the eigen values on the main diagonal. Furthermore, let choose the following channel matrix $\tilde{H} = WQ^T$ where $W \in \mathbb{R}^{n_r\times n_t}$ is diagonal matrix with unit elements on its diagonal, therefore $\tilde{H}$ is full rank with orthonormal rows. A lower bound for $\sup_H C^*(H,\mathbf{X}')$ is given by

$$\begin{aligned}
\sup_H C^*(H,\mathbf{X}') &\ge C^*(\tilde{H},\mathbf{X}') && (16)\\
&= \sup_{\mathbf{Z}:E||\mathbf{Z}||^2\le P} I(\tilde{H}\mathbf{Z}; \tilde{H}\mathbf{Z}+\tilde{H}\mathbf{X}') && (17)\\
&= \sup_{\mathbf{Z}:E||\mathbf{Z}||^2\le P} I(\tilde{H}\mathbf{Z}; \tilde{H}\mathbf{Z}+\tilde{\mathbf{X}}) && (18)\\
&\ge I(\tilde{\mathbf{Z}}; \tilde{\mathbf{Z}}+\tilde{\mathbf{X}}) && (19)\\
&= \sum_{i=1}^{n_r} I(\tilde{Z}_i; \tilde{Z}_i+\tilde{X}_i) && (20)\\
&\ge \frac{1}{2}\sum_{i=1}^{n_r}\log_2\left(1+\frac{P/n_r}{\sigma^2_{\tilde{X}_i}}\right) && (21)\\
&= \frac{n_r}{2}\sum_{i=1}^{n_r}\frac{1}{n_r}\log_2\left(1+\frac{P/n_r}{\sigma^2_{\tilde{X}_i}}\right) && (22)\\
&\ge \frac{n_r}{2}\log_2\left(1+\frac{P/n_r}{\frac{1}{n_r}\sum_{i=1}^{n_r}\sigma^2_{\tilde{X}_i}}\right) && (23)\\
&\ge \frac{n_r}{2}\log_2\left(1+\frac{P/n_r}{\frac{1}{n_r}P\frac{n_r}{n_t}}\right) && (24)\\
&= \frac{n_r}{2}\log_2\left(1+\frac{n_t}{n_r}\right), && (25)
\end{aligned}$$

where in (18) we define $\tilde{\mathbf{X}} = \tilde{H}\mathbf{X}'$, thus $R_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} = W\Lambda_{\mathbf{x}'}W^T$ which is $n_r \times n_r$ diagonal matrix where $R_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}(ii) = \Lambda_{\mathbf{X}'}(ii)$ for $\mathrm{i} = 1\ldots\mathrm{n_r}$. For $\tilde{\mathbf{Z}} \sim \mathcal{N}(0,\frac{P}{n_r})$ i.i.d Gaussian random vector we have (19), (20) is by using chain rule for mutual information. While (21) is from the fact that for additive channel $I(\tilde{Z}_i; \tilde{Z}_i+\tilde{X}_i) \ge I(\tilde{Z}_i; \tilde{Z}_i+\tilde{X^G}_i) = \frac{1}{2}\log_2(1+\frac{P/n_r}{\sigma^2_{\tilde{\mathbf{x}}}})$ where $\tilde{X^G}_i \sim \mathcal{N}(0,\sigma^2_{\tilde{\mathbf{x}}})$, i.e, the additive channel with Gaussian noise has the lowest capacity.

Since $\log(1+1/x)$ is a concave w.r.t x, (24) is due Jensen Inequality, while (25) is due to $\sum_{i=1}^{n_r}\sigma^2_{\tilde{\mathbf{x}}} \le P\frac{n_r}{n_t}$, since $R_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}$ has the lowest n_r eigen values of $R_{\mathbf{x}'\mathbf{x}'}$ where $tr(R_{\mathbf{x}'\mathbf{x}'}) \le P$.

On the other hand, we show that

$$\sup_H C^*(H, \mathbf{X}^*) \le \frac{n_r}{2}\log_2(1+\frac{n_r}{n_t}). \tag{26}$$

For $\mathbf{X}' = \mathbf{X}^*$ (4) becomes

$$C^*(H, \mathbf{X}^*) = \sup_{\mathbf{Z}:E||\mathbf{Z}||^2\le P} I(H\mathbf{Z}; H\mathbf{Z} + H\mathbf{X}^*), \tag{27}$$

since $\mathbf{X}^* \sim \mathcal{N}(0, \frac{P}{n_t}I_{n_t})$ the optimum is achieved by $\mathbf{Z}$ with Gaussian distribution [1, p.488], thus

$$\begin{aligned} C^*(H, \mathbf{X}^*) = \quad & \sup \left\{ \frac{1}{2}\log_2 \frac{|HR_{\mathbf{zz}}H^T + \frac{P}{n_t}HH^T|}{|\frac{P}{n_t}HH^T|} \right\}. \\ s.t: \quad & R_{\mathbf{zz}} \succeq 0 \\ & tr(R_{\mathbf{zz}}) \le P. \end{aligned} \tag{28}$$

Let define the matrix $D \triangleq HR_{\mathbf{zz}}H^T + \frac{P}{n_t}I_{n_r}$. The rows of H are orthonormal then $HH^T = I_{n_r}$, hence the denominator in (28) is constant, while the nominator is:

$$\begin{aligned} |HR_{\mathbf{zz}}H^T + \frac{P}{n_t}I_{n_r}| &\le \prod_{i=i}^{n_r} D_{ii} && (29)\\ &\le \left(\frac{1}{n_r}\sum_{i=1}^{n_r} D_{ii}\right)^{n_r} && (30)\\ &= \left(\frac{tr(HR_{\mathbf{zz}}H^T) + \frac{P}{n_t}n_r}{n_r}\right)^{n_r} && (31)\\ &\le \left(\frac{P + P\frac{n_r}{n_t}}{n_r}\right)^{n_r} && (32)\\ &= \left(\frac{P}{n_r} + \frac{P}{n_t}\right)^{n_r}, && (33) \end{aligned}$$

where (29) is from the Hadamard Inequality [1, p.502], (30) is due to the Arithmetic-Geometric Mean Inequality, (32) is from the fact that $tr(HR_{\mathbf{zz}}H^T) \le P$, since H has orthonormal rows. Using (28) and (33), $C^*(H, \mathbf{X}^*)$ can be written as

$$\begin{aligned} C^*(H, \mathbf{X}^*) &\le \frac{1}{2}\log_2 \left(\frac{\frac{P}{n_r} + \frac{P}{n_t}}{\frac{P}{n_t}}\right)^{n_r} && (34)\\ &= \frac{n_r}{2}\log_2\left(1 + \frac{n_t}{n_r}\right). && (35) \end{aligned}$$

It can be noticed that equality is achieved for $R_{\mathbf{zz}} = \frac{P}{n_r}H^T H$, since (28) is a convex problem over convex constraints this solution is a global maximum. Finally, (26) is proved since (33) holds for any channel matrix H especially for the worst channel matrix H, thus

$$\sup_H C^*(H, \mathbf{X}^*) \leq \frac{n_r}{2} \log_2 \left(1 + \frac{n_t}{n_r}\right),$$

with equality if $R_{\mathbf{zz}} = \frac{P}{n_r} H^T H$.

From (26) and (15) we have that

$$\sup_H C^*(H, \mathbf{X}^*) \leq \sup_H C^*(H, \mathbf{X}'), \ \forall R_{\mathbf{x}'\mathbf{x}'}. \tag{36}$$

The proof follows since $\mathbf{X}' = \mathbf{X}^*$ minimizes $\sup_H C^*(H, \mathbf{X}')$.

B. Proof of Theorem 1

Using (5) C^* is given by

$$\begin{aligned} C^* &= \inf_{E||\mathbf{X}'||^2 \leq P} \sup_H C^*(H, \mathbf{X}') & (37)\\ &= C^*(H, \mathbf{X}^*) & (38)\\ &\leq \frac{n_r}{2} \log_2(1 + \frac{n_t}{n_r}), & (39) \end{aligned}$$

where (38) is from the definition in (4), (39) is due to (35).

Now, let us derive the bound for $n_r \geq n_t$. In Lemma 2 it was shown that without loss of generality the problem can be reduced to $n_t \times n_t$ full rank orthogonal channel matrix, therefore the capacity loss is bounded by $C^* \leq \frac{n_r}{2} \log_2(1 + \frac{n_t}{n_r}) = \frac{n_t}{2}$.

C. Proof of Theorem 2

We assume a MIMO channel $Y = H\mathbf{X} + N$, where the noise $N \sim \mathcal{N}(0, I_{n_r})$ without loss of generality, since for a given channel matrix and additive Gaussian noise statistics, the problem is equivalent to a new channel matrix and uncorrelated Gaussian noise with unit variance.

The channel capacity can be calculated using Singular Value Decomposition (SVD) of H, thus

$$H = Q_2 \Delta Q_1^T, \tag{40}$$

where $Q_2 \in \mathbb{R}^{n_r \times n_r}$, $Q_1 \in \mathbb{R}^{n_t \times n_t}$ are unitary matrices and $\Delta \in \mathbb{R}^{n_r \times n_t}$ has zero elements off-diagonal. Using unitary transformations $\tilde{\mathbf{Y}} = \mathbf{Q_2^T Y}$ and $\mathbf{X} = Q_1 \tilde{\mathbf{X}}$ at the encoder and decoder, respectively, the capacity is given by

$$C = \max_{E||\mathbf{X}||^2 \leq P} I(\mathbf{X}; H\mathbf{X} + \mathbf{N}) = \max_{E||\tilde{\mathbf{X}}||^2 \leq P} I(\tilde{\mathbf{X}}; \Delta\tilde{\mathbf{X}} + \tilde{\mathbf{N}}), \tag{41}$$

where $\tilde{\mathbf{N}} = Q_2^T \mathbf{N}$, hence $\tilde{\mathbf{N}} \sim \mathcal{N}(0, I_{n_r})$, while the power constraint is $E||\mathbf{X}||^2 = E||\tilde{\mathbf{X}}||^2 \leq P$.
Generally, for any channel matrix with $rank(H) = r$ the problem can be reduced to $r \times r$ square problem, since Δ has only r non-zero elements on the diagonal (assuming that the non-zero elements are Δ_{ii} for $i = 1 \ldots r$), thus the capacity is given by

$$C = \max_{E||\mathbf{X}'||^2 \leq P} I(\mathbf{X}'; R\mathbf{X}' + \mathbf{N}'), \tag{42}$$

where

$$\begin{aligned} R=\{R_{ij}\} &= \Delta_{ij}, i,j=1\dots r \\ X'_i &= \tilde{X}_i, i=1\dots r \\ N'_i &= \tilde{N}_i, i=1\dots r \end{aligned}$$

where $\mathbf{N}' \sim N(0, I_r)$ and R is diagonal matrix with non zero elements on the diagonal. This model corresponds to a parallel channels model with colored noise by multiplying the receiver input by R^{-1}. The capacity of parallel channels is achieved using water filling optimization [1], the noise covariance matrix $\Lambda = diag(\lambda_1 \dots \lambda_r)$ is diagonal matrix where $\lambda_i = \frac{1}{R_{ii}^2} \geq 0$, thus the capacity is

$$\begin{aligned} C &= \frac{1}{2}\sum_{i=1}^{r} \log_2(1+\frac{(\nu-\lambda_i)^+}{\lambda_i}) \\ s.t: & \quad \sum_{i=1}^{r}(\nu-\lambda_i)^+ \leq P. \end{aligned} \tag{43}$$

On the other hand, the rate achieved using i.i.d Gaussian input $\mathbf{X}^* \sim \mathcal{N}(0, \frac{P}{n_t} I_{n_t})$ is given by

$$\begin{aligned} I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}) &= \frac{1}{2}\log_2 \frac{|H\frac{p}{n_t}I_{n_t}H^T + I_{n_r}|}{|I_{n_r}|} && (44) \\ &= \frac{1}{2}\log_2 |\frac{p}{n_t}HH^T + I_{n_r}| && (45) \\ &= \frac{1}{2}\log_2 |\frac{p}{n_t}Q_2\Delta\Delta^T Q_2^T + I_{n_r}| && (46) \\ &= \frac{1}{2}\log_2 |Q_2(\frac{p}{n_t}\Delta\Delta^T + I_{n_r})Q_2^T| && (47) \\ &= \frac{1}{2}\log_2 |\frac{p}{n_t}\Delta\Delta^T + I_{n_r})|, && (48) \end{aligned}$$

where (46) is from (40). Since $\Delta\Delta^T = R^2$, we have that

$$I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}) = \frac{1}{2}\log_2 \prod_{i=1}^{r}(1+\frac{P}{n_t}R_{ii}^2). \tag{49}$$

Since $R_{ii}^2 = \frac{1}{\lambda_i}$ for $i = 1\dots r$, it can be written as

$$I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}) = \frac{1}{2}\sum_{i=1}^{r}\log_2 \frac{\frac{P}{n_t}+\lambda_i}{\lambda_i}. \tag{50}$$

Finally, the capacity loss is the difference between (43) and (50), therefore

$$\begin{aligned} C - I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}) &= \frac{1}{2}\sum_{i=1}^{r}\left\{\log_2(\frac{\lambda_i + (\nu-\lambda_i)}{\lambda_i}) - \log_2 \frac{\frac{P}{n_t}+\lambda_i}{\lambda_i}\right\} && (51) \\ &= \frac{1}{2}\sum_{i=1}^{r}\left\{\log_2(\frac{\max(\lambda_i,\nu)}{\lambda_i}\cdot\frac{\lambda_i}{\lambda_i+\frac{P}{n_t}})\right\} && (52) \\ &= \frac{1}{2}\sum_{i=1}^{r}\left\{\log_2\frac{\max(\lambda_i,\nu)}{\lambda_i+\frac{P}{n_t}}\right\}. && (53) \end{aligned}$$

Let define $\mathcal{I} \triangleq \{i : \nu - \lambda_i > P/n_t, i = 1, \ldots, r\}$, where $|\mathcal{I}|$ is the cardinality of $\mathcal{I}$, thus

$$\begin{aligned}
C - I(\mathbf{X}^*; H\mathbf{X}^* + \mathbf{N}) = \ \leq &\ \frac{1}{2}\sum_{i\in\mathcal{I}} \log_2 \frac{\nu}{\lambda_i + \frac{P}{n_t}} && (54)\\
\leq &\ \frac{1}{2}\sum_{i\in\mathcal{I}} \log_2 \frac{\nu - \lambda_i}{\frac{P}{n_t}} && (55)\\
= &\ \frac{1}{2}|\mathcal{I}| \cdot \sum_{i\in\mathcal{I}} \frac{1}{|\mathcal{I}|} \log_2 \frac{\nu - \lambda_i}{\frac{P}{n_t}} && (56)\\
\leq &\ \frac{1}{2}|\mathcal{I}| \cdot \log_2 \Big(\frac{1}{|\mathcal{I}|} \sum_{i\in\mathcal{I}} \frac{\nu - \lambda_i}{\frac{P}{n_t}} \Big) && (57)\\
\leq &\ \frac{1}{2}|\mathcal{I}| \cdot \log_2 \frac{n_t}{|\mathcal{I}|} && (58)\\
= &\ \frac{n_t}{2} \cdot \frac{|\mathcal{I}|}{n_t} \cdot \log_2 \frac{n_t}{|\mathcal{I}|}, && (59)
\end{aligned}$$

where (55) is from the fact that ν, $\frac{P}{n_t}$ and λ_i are positive, (57) is from Jensen's inequality, (58) follows from the power constraint of (43).
The function $\frac{|\mathcal{I}|}{n_t} \cdot \log_2 \frac{n_t}{|\mathcal{I}|}$ is concave function w.r.t $|\mathcal{I}|$ and has global maximum at $|\mathcal{I}| = \frac{n_t}{e}$, which implies that we can split the bound into two cases:
For $0 \leq n_r < \frac{n_t}{e}$ - we have that $|\mathcal{I}| \ \leq r \leq n_r$, since the function $\frac{|\mathcal{I}|}{n_t} \cdot \log_2 \frac{n_t}{|\mathcal{I}|}$ is increasing over the interval $|\mathcal{I}| \in [0, n_t/e)$, therefore the capacity loss is bounded for $|\mathcal{I}| = n_r$ by

$$\begin{aligned}
C_G^* &= \frac{n_t}{2} \cdot \frac{n_r}{n_t} \log_2(\frac{n_t}{n_r}) && (60)\\
&= \frac{n_r}{2} \cdot \log_2(\frac{n_t}{n_r}). && (61)
\end{aligned}$$

For $n_r \geq \frac{n_t}{e}$ - the capacity loss is bounded by the loss at the global maximum, thus

$$C_G^* = \frac{n_t}{2} \cdot \frac{\log_2(e)}{e}. \qquad (62)$$

References

[1] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[2] I. Csiszar and J. Korner, *Information theory - coding theorems for discrete memoryless systems*, Academic Press, New York, 1981.

[3] D. Hosli and A. Lapidoth, *How good is an isotropic Gaussian input on a MIMO Ricean channel?*, Presented at the ISIT 2004, Chicago, USA (Jul. 2004), 291.

[4] D. P. Palmoar, J. M. Cioffi, and M.A. Lagunas, *Uniform power allocation in MIMO channels: A game-theortic approach*, IEEE Trans. Information Theory **IT-49** (Jul. 2003), 1707–1727.

[5] N. Shulman and M. Feder, *Static broadcasting and common information broadcasting in general*, Presented at the ISIT 2000, Sorrento, Italy (Jun. 2000).

[6] R. Zamir and U. Erez, *A Gaussian input is not too bad*, IEEE Trans. Information Theory **IT-50** (Jun. 2004), 1362–1367.